



UN
Big Data
Hackathon

WEBINAR

UN Big Data Hackathon
Pre-Hackathon Webinar

The 2022 UN Big Data Hackathon in numbers...



4 days



60+ countries



450+ teams



**1000+
participants**



2022 Theme

Using Big Data and Data Science to develop ideas and solutions to address Global Challenges and help achieve Sustainable Development Goals; notably to support policies caused by:

- The disruption to Global Value Chains and Economic Globalization due to disasters, conflicts, restrictions, blockages
- The impact of Climate Change on society as part of monitoring SDG 13
- The rise of food and energy prices affecting vicious cycles of poverty, hunger, and inequalities

Platform introduction: AWS



Thuan Tran

Senior Solutions Architect



Kathleen McGeer

Senior Customer Solutions Manager



Raju Rangan

Senior Solutions Architect



Jordan Robert

Account Manager

What should I do during the hackathon? (1/3)

Examples of guiding questions

- Given disasters, conflict/war, restrictions, and blockage, how can big data be utilized to measure the impact of such disruption? For example, what was the cost to the shipping companies/customers when the Suez canal was blocked?
- What are the impacts of the drying up of rivers around the world (i.e., Yangtze River, Colorado River, Rhine River) on the economy, environment, and society?
- How to monitor vulnerable countries that depend on imports for their food security considering the increase in prices and supply limit due to widespread restrictions on export?

What should I do during the hackathon? (2/3)

Deliverables

Hackathon = hack your way to an analytical solution that addresses the impact on SDGs of disasters, climate change and/or rise of energy/food prices

A **reference database** with various datasets related to the theme will be provided. **You can also use any public dataset even if it is not provided in the database.**

In line with the theme, you should leverage data analysis/big data to develop the prototype of a data product such as:

- Analytical reports
- Interactive dashboards with enhanced visuals
- Advanced Machine Learning models

What should I do during the hackathon? (3/3)

Deliverables

Teams must submit:

1. Presentation explaining their solution (free format).
2. Video with maximum length of 10 minutes = voice-over of the presentation
3. Coding scripts.

If one of these deliverables is missing, the team will get a penalty

Big Data Experts Track

Examples of outputs

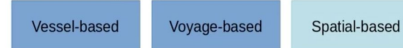
Example 1: BlueCarbon, overall winning team of the 2020 AIS Hackathon

Using the AIS data provided by the UN Global Pulse platform, the team geographically distributed CO₂ emissions from shipping based on individual vessel locations and activity using a Machine Learning Model. They subsequently developed an interactive dashboard to map these emission distributions for different time periods.

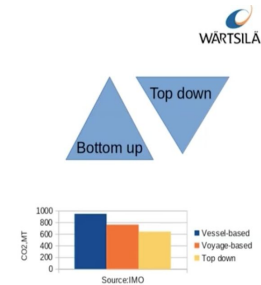


EMISSIONS ESTIMATION: APPROACHES AND DEVELOPMENT

Methodologies:
(IMO GHG Study 2020)



Validation (Bottom-up / Top down / MRV)
Carbon intensity metrics: T, T/DIST, T/TIME, T/DWT/DIST
Accuracy (missing data)
Average characteristics and dependencies



Model development:



Reference: BlueCarbon's presentation video

<https://www.youtube.com/watch?v=qmybyzV5R8A>



Big Data Experts Track

Examples of outputs

Example 2: DogCat, student winning team of the 2020 AIS Hackathon

Using the AIS data provided by the UN Global Pulse platform, the team analyzed the impact of COVID-19 on three major sectors: commodities, bulk carriers and trade countries

Topic : The difference in the impact of COVID-19 to container shipping

Motivation

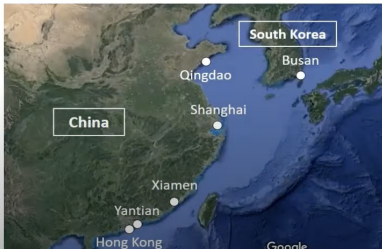
Is there any difference in the impact of COVID-19?
e.g.) Deployed vessels becomes smaller by the low demand

Objective

To measure the difference in the impact of COVID-19 to several factors

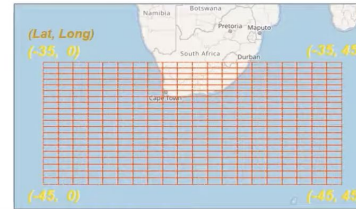
Target

Data : Arrival data on AIS data
Ship : Container ship
Port : Busan, Qingdao, Yantian, Shanghai, Xiamen, Hong Kong, Los Angeles
Term : 2019/01/08 - 04/23
2020/01/08 - 04/22
Factor : Shipping Alliance and Vessel size



Research Policy

- ① Set up a polygon area around the Cape of Good Hope and generate meshes in the area. (20 × 20 split)



- ② Extraction of speed and position data for each vessel in the area.
- ③ Using AIS data and other ship movement data, GHG emissions are calculated for each mesh and individual ship and the total values are tabulated.

Reference: DogCat's presentation video

https://www.youtube.com/watch?v=X_jmNy4qbjE



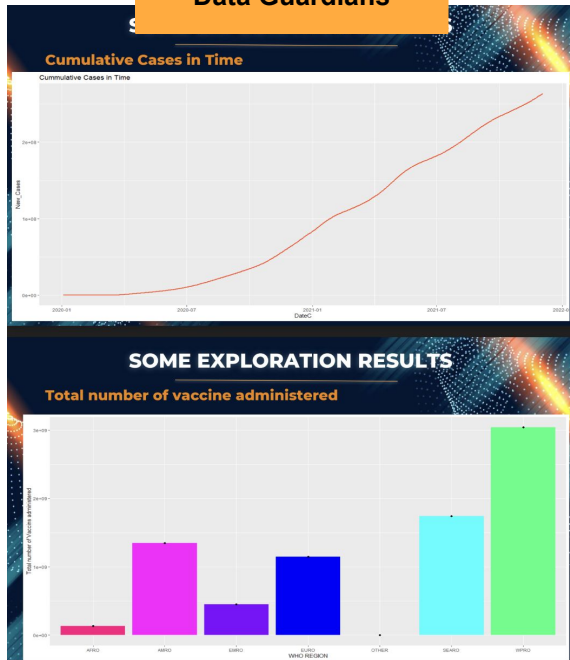
Youth Track

Examples of outputs

Example 1:

Reports providing detailed statistical analysis or data correlations between different datasets

Data Guardians



UN Youth Hackathon projects (sample):

<https://drive.google.com/drive/folders/1MxLUONvDnD224fvHPCzKUatx-s3OJ-T7>

Section III

Most of the industries were shutdown to lower mobility. There have been negative economic implications due to restrictions and decreased travel business worldwide.

The negative economic implications may increase inequality, discrimination and medium and long-term unemployment if not properly addressed by right policies.

CAS-ZJSO

Most countries are currently trying to combat the virus spread by screening for COVID-19 in large numbers and maintaining strict quarantines, entry bans policies, as well as other limitations for citizens in or recent travelers to several countries in the most affected areas.

Unemployment rate yearly by global and regions

As a key performance indicator of economic health and a crucial metric for judging policy outcomes, unemployment rate typically calculated on a monthly, quarterly or annual basis.

The ILO provides unemployment rate data covering most of countries and areas until 2021. After the 2030 agenda for sustainable development was adopted by world leaders in 2015, the global yearly unemployment rate constantly decrease to 5.4% in 2019.

The yearly unemployment rate in G20 and G7 countries decreased by 0.6% and 2.3% respectively from the peak in 2011 to 5.5% and 4.2% respectively in 2019. At the SDG Summit held in September 2019, SDG Member States renewed international commitment to fulfil the 2030 agenda for sustainable development.

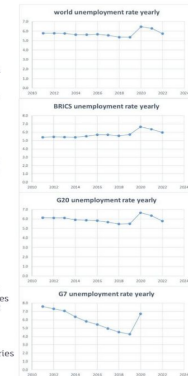
Yet, in only of brief period of time, the COVID-19 has disrupted efforts to achieve the 2030 agenda for sustainable development.

The COVID-19 pandemic has upended the global labor market, with massive job losses and spike in unemployment to its highest level.

The yearly global unemployment rate increased by 1.1% to 6.5% in 2020. The dramatic increase in BRICS, G20, G7 countries by 1.0%, 1.2% and 2.4% respectively to 6.4%, 6.7% and 6.7%.

Thanks to strong policy support, accelerated vaccine rollout and the application of digital solutions to maintain business and consumption, growth resumed forcefully in the world in 2021.

The unemployment rate dropping to 5.5%, showed a continued recovery from 2020 to 2021 after the COVID-19 pandemic. The BRICS and G20 countries also have the same economic recovery trends.



Youth Track

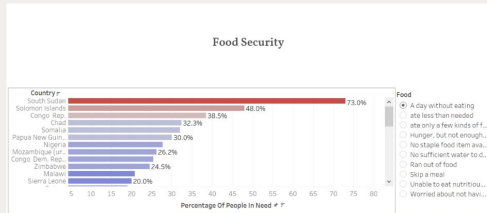
Examples of outputs

UN Youth Hackathon projects (sample):
<https://drive.google.com/drive/folders/1MxLU0NvDnD224fVHPcZKUatx-s3OJ-T7>

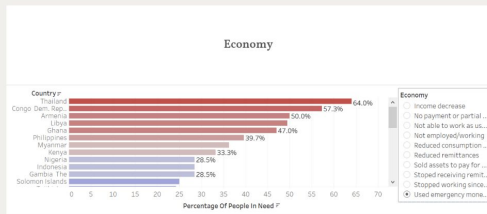
Example 2:
Interactive dashboard with enhanced analytical visuals

Team IdeaX

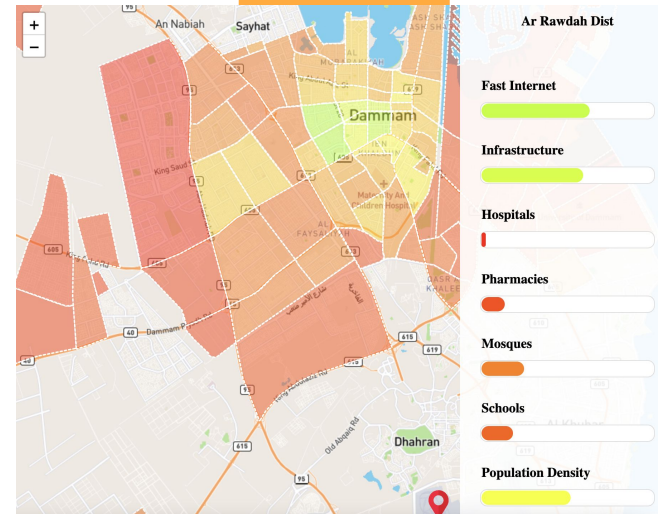
visualization



visualization



Doves



Youth Track

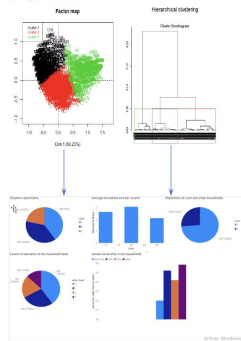
Examples of outputs

Example 3: Advanced Machine Learning Models

Team Sustainability, overall winner

Conclusion: Our solution

A statistical segmentation to better understand the impact of a household socio-economic characteristics on their vulnerability to COVID-19 and their consequences.



A integrated prediction model in order to assess the vulnerability of households to COVID-19 regarding their income, food security and education access



UN Youth Hackathon projects (sample):
<https://drive.google.com/drive/folders/1MxLU0NvDnD224fVHPcZKUatx-s3OJ-T7>

Data Rockstars

Pandemic Score Methodology Overview

1. Researching Data 2. Preprocessing Data 3. Creating Factors

- Look for data specifically for the pandemic period.
 - Education Survey
 - Time spent at home
- Survey data: questions and answers were converted into indicators and flags.
- Each factor is built upon some metrics derived from the available variables. The final factor is an equally weighted average over these metrics scaled to the maximum points attributed to the main category.

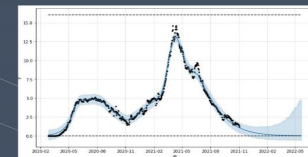
What does the score components measure?

- Online Learning**
 - Adherence
 - Effectiveness assessment
 - Inclusion
 - Support to teachers
 - Engagement
- Learning**
 - Adjustments to curriculum and calendar.
 - Learning gaps
 - Learning losses
- School Closures**
 - Total time of full closure
 - Current state of full closures
- School Reopening**
 - Adherence
 - Strategies
 - Support to staff
- Vulnerable groups**
 - Special support for distance learning and reopening
- Tests and Vaccines**
 - Rate of people fully vaccinated
 - Rate of tests
 - Rate of vaccination
- Mobility**
 - Relative change of time spent at home relative to the pre-pandemic period.
- Forecasting**
 - Relative change in new cases and deaths (forecasted) relative to the previous period (2 weeks window)

Forecasting Methodology

To predict COVID-19 new cases and deaths, we applied an additive model with non-linear trends fit daily. This approach is robust to missing data and shifts in the trend.

Algorithm: Prophet
Y: Smoothed new cases and deaths per million
Training data: JAN/2020 - OCT/2021
Validation: NOV/2021
Out of time prediction: JAN/2022
Mean Squared Error: 8.65



Forecasting example for Brazil's new cases per million



UNBDH Timeline

Join the [mailing list](#) to be updated (webinars, deadlines, winners!)



- Access to the databases will be granted starting from **November 8th at 2:00 PM Yogyakarta Time (UTC+7)**.
- Proposals must be submitted on **November 11th at 2:00 PM Yogyakarta Time (UTC+7)**.



Earth Hacks Responsible Innovation Workshop

The workshop will follow the themes of responsible innovation, innovation and the climate crisis, *technosolutionism*, and innovation culture.

Designed to help teams create a better solution in line with the theme.



The banner features a teal header with the 'Earth Hacks' logo (a lightbulb with a leaf inside) and the 'UN Big Data Hackathon' logo (a blue and green geometric shape). The main content is on a yellow background with a central illustration of a globe with orbits and stars. The title 'RESPONSIBLE INNOVATION IN THE CLIMATE CRISIS' is written in large, bold, black serif font, and the date 'November 4, 8am EST' is centered below it.

Earth Hacks

UN Big Data Hackathon



**RESPONSIBLE
INNOVATION IN THE
CLIMATE CRISIS**

November 4, 8am
EST

Communication Platform >> SLACK

General channels

- Channel **“announcements”**: countdown, important information...
- Channel **“data”**: share useful data insights or link to newly found public datasets.
- Find the channel **“technical-issues”** to seek technical support related to platform access
- Ask any questions that doesn't go into the other categories in the **“submission-questions”**
- Relax and get to know other participants in the **“lounge”** channel

Regional channels

- You will be added to a private regional channel to get to know teams from your region

Big data experts channel

- Big data experts can communicate and share insights on this channel

Teams can create a private channel to have a collaborative work.

**Teams can support each other through Slack at any time. The UNBDH Mentors from different regions will support teams on SLACK and will be available between 8:00 AM and 6:00 PM for each respective time zone.*

Platform introduction: UNGP AIS



Amna Gul

Data Scientist
Asian Development Bank (ADB)



Sean Lovell

Information Systems Officer
United Nations Statistics Division

Platform introduction: ArcGIS



Brian Baldwin

Senior Solution Engineer at Esri

Judging Criteria

Criterion	Description	Weight
Theme	Does the team develop a solution that is in line with the event's theme? Is the solution realistic and scalable?	30%
Innovation	Does the team's idea show "Out of the Box" thinking? Is the team's solution groundbreaking?	30%
Methods	Is the technology behind the idea impressive? Does the solution use any new methods?	20%
Presentation*	How well was the project presented? Does it make the idea more appealing?	10%
Visualization*	How well has the report's data visualization been done?	10%

* This criteria is for separate prizes Best presentation and best visualization

Penalty

- Did the team submit codes, slides, video?
- Did the team use private data in the project?
- Did the big data expert tract team use big data?

Last advice

1. Have a good communication strategy with your team!
2. Start populating the slides as you go; don't leave it for the last day!
3. Manage your time carefully: 4 days is a short amount of time
4. Divide and conquer: each team member shall work on different tasks and sync regularly so you can get the best results
5. If you're stuck, don't stop: reach out to us and mentors on SLACK or by email! Finishing is better than abandoning :)



Q&A

Do you have additional questions?

un-big-data-hackathon@unmgcy.org

Follow us on:

<https://unstats.un.org/bigdata/events/2022/hackathon/>



[@unbigdatahackathon](https://twitter.com/unbigdatahackathon)



[@unbigdatahackathon](https://www.instagram.com/unbigdatahackathon)



[@unbigdatahackathon](https://www.facebook.com/unbigdatahackathon)

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

